

TP DATAWAREHOUSE

Le but du TP est d'intégrer des données provenant de fichiers dans votre base de données Oracle en utilisant TALEND ; Talend est un ETL ; on va donc faire de l'Extraction, Transformation Load (chargement).

Ouvrez un terminal et lancez la commande suivante : `/opt/Studio/TOS_ESB-linux-gtk-x86_64`

Elle vous permettra de lancer Talend sur vos postes.

Vous trouverez le fichier `employees.csv` dans le dossier
`@\\iluvatar\export\documents\is3+4\sgbd\lekpa`.

1- Intégration des données d'un fichier csv dans une table Oracle

- a. Créez un nouveau Job ; à gauche faites un clic droit sur job Design → create job ;
- b. Créez une connexion à votre base de données oracle dans laquelle les données vont être intégrées. Pour cela, allez dans « metadata » puis clic droit sur « db connexion ». Nommer votre connexion et renseigner les informations sur cette dernière. Puis cliquer sur next. Choisir « Oracle with SID » comme type de connexion et renseigner les informations demandées.
- c. Les fichiers à intégrer sont dans le dossier ... copier les différents fichiers et stocker les en local.
- d. Créer une connexion au fichier xxx en créant un « File delimited ». Remplissez les différents champs ; renseigner la location du fichier csv. Ouvrir le fichier pour vérifier le délimiteur des données afin de choisir le bon délimiteur.
- e. Cliquez sur le job pour ouvrir l'interface du job. Portez et déposez la connexion au fichier csv et choisissez « tfileInputDelimited » pour indiquer que le fichier csv est votre source.
- f. Portez et déposez la connexion à votre base de données Oracle et choisissez « tOracleOutput » pour désigner que la base Oracle est votre destination.
- g. Créer un lien entre votre source et votre destination : clic droit sur filedelimited -> row->main puis cliquez sur oracle connexion.
- h. Cliquez sur la connexion à la base de données Oracle et regardez la section « basic settings ». Vérifiez que vous avez bien renseigné le nom de la table. Et vérifiez « action on table » et « action on data »
- i. Lancer votre job ; connectez vous à sql developer et vérifiez que la table a bien été créée et que les données y ont été insérées.

2- Utilisation de Tmap

- a. Regarder en ligne à quoi sert le composant TMAP de Talend.
- b. Utiliser le composant TMAP pour séparer les données contenues dans le fichier csv. Une première table qui va contenir les informations de l'employé

(firstname, lastname, email, gender, phone, department, job title). Une deuxième table qui va contenir (le salaire et le 'nom+prenom' de l'employé ; nom+prenom étant unique). Pour cela, il faudra créer deux sorties dans le composant TMAP.

- c. Faite une jointure entre les tables employees et department.
- d. Utiliser le composant TMAP pour masquer les deux premiers chiffres et les deux derniers chiffres de numéro de téléphone des employés